Alberta Assessment of School Mathematics

EXECUTIVE SUMMARY

APRIL, 1979

Alberta

Minister's Advisory Committee on Student Achievement

Executive Summary

of a report

on

Alberta Assessment of School Mathematics

A Study Conducted for

The Minister's Advisory Committee on Student Achievement

by

A.T. Olson, D. Sawada, S.E. Sigurdson

The University of Alberta

under contract to the Planning and Research Branch

of Alberta Education

1979

ABSTRACT

Students at each of grades 3, 6, 9, and 12 responded to 120 questions. About 70 percent of the questions were multiple choice; the remainder were open-ended. Between 54 and 92 percent of the 120 questions were appropriate to the target grade, and the others were review items appropriate to lower grades or preview questions suitable for higher grades. Review and preview items were administered as common items in several grades to provide cross-sectional information.

On May 17, 1978, the tests were administered to a 10 percent random sample of schools consisting of 102 schools offering grade 3; 101 schools offering grade 6; 56 schools offering grade 9; and 50 schools offering grade 12. A form of matrix sampling was employed so that each student did not respond to all 120 test items; instead, different sets of test exercies were randomly assigned to each student.

Performance in number facts and computation (adding, subtracting, multiplying and dividing) was satisfactory at all grade levels, but performances in problem solving, geometry, measurement and consumer mathematics were unsatisfactory. The difficulty with problem solving might be attributed to students' reading abilities as well as to the complex reasoning skills required, but no firm evidence is available to support this speculation. Weaknesses in geometry, measurement and consumer mathematics might be related to the emphasis these areas receive in the Alberta mathematics curriculum. Geometry studies at the elementary level have only recently been increased, the metric system of measurement is relatively new to Alberta school programs, and consumer mathematics, although considered an important skill for daily living, receives little or no attention in the senior high school.

This study found that boys performed better than girls on all aspects of mathematics and this difference became more pronounced as the grade level increased. School size also showed a relationship to achievement differences, and interestingly, the differences interacted with grade levels. Grade 3 students in small schools performed best; in grade 6 the medium-sized school began to emerge as dominant; for grade 9 the medium school showed clear dominance; and at the grade 12 level school size failed to show any significant differences in achievement. The large elementary school quite consistently had the lowest levels of performance. Regarding the type of high school mathematics program, students who had taken Math 30 had the highest scores. It is not surprising that the Math 30 students performed best in the areas where they had received the greatest amount of instruction (number work and algebra) and that there were fewer differences among students in different mathematics programs in areas where the amount of instruction is more equivalent (geometry, measurements, statistics and consumer mathematics).

FOREWORD

The Minister's Advisory Committee on Student Achievement (MACOSA) was established by ministerial order in October 1976 in response to growing concerns expressed by the public-at-large, government, labor, business, students and educators regarding the quality and standards of basic education in Alberta.

MACOSA commissioned a number of studies, primarily to provide basic information for a summary of current levels of achievement in Alberta and to provide baseline data for future assessment. These studies fell into three categories: (1) preliminary studies, (2) achievement studies, and (3) other studies.

This achievement study, Alberta Assessment of School Mathematics, was designed to provide information about current levels of achievement in mathematics among students in Alberta schools and to provide a data base for future assessments.

This report, which represents the findings and conclusions of the researchers, was presented to MACOSA as information.

ACKNOWLEDGEMENTS

Several people provided invaluable service during the conduct of the assessment. The contractors would like to take this opportunity to express their appreciation and thanks to the following people.

Teachers who participated in item validation and pilot testing:

Teacher	School
Darryl Smith	Austin O'Brien, Edmonton
Ken Auch	Goldbar, Edmonton
Marilyn Cameron	Kildare, Edmonton
Elizabeth Beaudry	J J Bowlen, Edmonton
Harry Johnson	Ottewell, Edmonton
Bruce Kabaroff	Jasper Place, Edmonton
Joy Bengert	St. Brides, Edmonton
Jim Hancheruk	Waverly, Edmonton
Francis MacLean	Percy Pegler, Okotoks
Mary Poffenroth	Red Deer Lake, Calgary
Ed Bagyinka	Senator Riley, High River
Larry Berg	Okotoks, Okotoks
Sylvia Hay	Percy Pegler, Okotoks
Dale Backlin	Spitzee, High River
Mary Trainor	Okotoks, Okotoks
Eugene Deputat	Oilfields, Black Diamond
Marie Romanko	Albert Lacombe, St. Albert
Roy Osland	Kildare, Edmonton
Fern Valliere	Albert Lacombe, St. Albert
Lucy Beauchemin	Albert Lacombe, St. Albert
Peter Holt	Albert Lacombe, St. Albert
Dan Tessari	Stratford, Edmonton
Ernest Mercer	Ottewell, Edmonton

Tom Maguire who was consulted on the design of the study.

David Robitaille, a member of the contract team for the British Columbia Mathematics Assessment, who consulted with us for two days.

John LeBlanc, a representative of the National Council Teachers of Mathematics on the National Assessment of Educational Progress (NAEP) who consulted with us in regards to the 1978-79 NAEP Mathematics Assessment.

The 13,079 students who took the time to write the test booklets and the teachers and principals who supervised the data collection.

Anita, Barbara, and Corrine, who typed thousands of pages.

Dan Precht of Computer Services, who programmed the data analysis.

And, finally, the members of the MACOSA Steering Committee: Clarence Rhodes (Chairman), Planning and Research Branch, Alberta Education; Al Anderson, Mathematics Consultant, Alberta Education; Al Gibb, Faculty of Education, University of Calgary; and Earle Warnica, Mathematics-Science Coordinator, Lethbridge Public School District. Harry Sherk, Executive Secretary and Project Director of MACOSA, was an ex officio member, and Morris Treasure of the Planning and Research Branch of Alberta Education served as recorder and consultant to the steering committee.

PREFACE

In doing the study reported herein the contractors took the position that the assessment of student achievement needs drastic re-conceptualization if the data gathered are to have any substantial and ongoing impact on educational programs and policies. The traditional "tests and measurement" point of view has not provided enough punch to be taken seriously by those who are involved in top level educational decision making. On the other hand, those involved with top level as well as low level decision making in regards to the economy do consult the Consumer Price Index (CPI) as a valid barometer of the Canadian economy. In fact the monthly fluctuations of the CPI are often treated as headline news. For these and other reasons the present study is reported as if a counterpart mathematics proficiency index (MPI) were already in existence. Terminology referring to various mathematical content indices is used throughout the report even though properly speaking the present study does NOT provide index numbers since in order to compute index numbers a baseline year is required for comparison. However, baseline data do not exist since the present study is the first of its kind to be cast as an index in Alberta. Nevertheless the contractors believe strongly that unless studies of the present kind are designed and interpreted as providing <u>indices</u>, their usefulness in educational decision making will be marginal at best. On this basis the contractors have taken liberties in interpreting the present results as providing estimates of index numbers even though only baseline values are provided. We request the indulgence of the readership in this regard.

I Introduction and Overview

1.1 Origin and Purpose

On January 25, 1977 the Minister's Advisory Committee on Student Achievement (MACOSA) through the Planning and Research Branch of Alberta Education issued a call for proposals that would deal with the assessment of current levels of mathematics skills in Alberta schools over four grade levels, 3, 6, 9, and 12. According to that document

> The purpose of the study will be to provide information concerning the current levels of mathematics skills and knowledge as demonstrated by students in Alberta schools and to provide a base for future assessments.

Just what type of information and in what form it was to be provided and the use to which it would be put was not clarified in the document. Apparently part of the task of the proposal writer was to specify just what type of information was required and what form that information should take in order to fulfill the purpose as stated.

1.2 The Problem

In viewing and reviewing the state of assessment in North America and in Europe, it became clear to the contractors that

1. the type of information that was desired had to be understandable and face valid,

2. the form of the information had to be similar to that provided for similar purposes in other aspects of human endeavor so that the meaning and interpretation of the information could be arrived at in relatively well known and familiar ways by anyone interested in mathematical competence.

Thus in the same way that the Consumer Price Index allows the layman as well as the economist to monitor the current cost of living, so too should the information provided by this study and subsequent studies, allow laymen and mathematics educators to monitor the current levels of mathematical competence of Alberta students.

In a nutshell, then, the problem of the study was defined as follows:

1. to provide a comprehensive and coherently structured set of index numbers that would faithfully reflect current levels of mathematical competence of Alberta students, and

2. to design a simple yet effective scheme whereby these index numbers could be updated on a regular basis in order that a valid monitoring of student mathematical competence could be provided through a set of indices.

In defining the problem in the above way, the traditional test item was cast in a new role and in new light: at one extreme an individual test item could serve to estimate a single index number; more likely, however, several test items would be combined in the form of arithmetic means to estimate index numbers for various aspects of mathematical competence. Under this interpretation, an individual test item such as "2 + 6 = ?" would have something in common with a loaf of bread: the test item contributes to the estimation of what might be called the Mathematical Proficiency Index just as the price of a loaf of bread contributes to the estimation of the Consumer Price Index. The notion of an index undergirds much of the current methodology of assessment and in particular forms the basis of the National Assessment of Educational Progress (NAEP) in the United States. It, as well, forms the conceptual and statistical basis for the present study. As such, much of the value of the current assessment will remain untapped if regular updatings of the index numbers are not provided in the near future, since only by obtaining a series of index numbers (over time) can one obtain indices by computing ratios with the index numbers provided by the current assessment used as the base.

1.3 Approach to the Problem

The major point of departure of this assessment is the use of the notion of an index as the proper statistical quantity to relay information regarding the status of student achievement. Without such an orientation, the results of this and other assessments like it will disappear into the archives leaving little or no impact on the educational world. With it, the continuity and base line comparisons provided should ensure the relevance and usefulness of the present results for now and in the future. The design of the present assessment was completed with the notion of an index given high priority.

The approach to the study is summarized in the following pages under five headings. More detailed specifications can be found in the Technical Report.

1.3.1 Survey Variables

The major variable was the set of index numbers to be developed in a structured and coherent manner. An individual index number was estimated by the percentage of respondents of a given

type who gave the correct answer on a particular test item (exercise). The structure and coherence of these estimates was provided by the use of two types of partitioning variables: (1) mathematical content specifications done according to level of thought process and "targeted" grade level within the Alberta curriculum, and (2) personological/demographic variables.

The mathematical content specification variables took the form of a 3-dimensional matrix with mathematics content (number, algebra, geometry, measurement, statistics, and consumer mathematics) forming the first dimension; level of thought process (Knowledge, Comprehension, and Application) forming the second dimension; and targeted grade level (Division I for grades 1-3, Division II for grades 4-6, Division III for grades 7-9 and Division IV for grades 10-12) forming the third dimension. The resulting matrix, called the Mathematics Content Specification Model, is shown in Figure I-1. For Divisions I, II and III the content domain was taken to be defined by the various curriculum guides published by Alberta Education for use in Alberta schools. For Division IV the content domain was taken to be reflective of the mathematics that would be useful in daily life as well as the various mathematics programs given in secondary schools.

The personological/demographic variables consisted of sex, grade level, school size, and at the high school level the type of mathematics program taken. These variables were included primarily to provide means of partitioning the percentage estimates into identifiable sources such as grade, sex and so on so that comparisons and contrasts could be made to aid in the interpretation of the indices. As well, the variable of grade level was used in a special way to provide a cross-sectional view of growth. More particularly, after a test item was selected according to the mathematical content specification model, a decision had to be made as to what grade level(s) the item would be given. If given at more than one grade level, the item was called a common item and was therefore useful for tracing growth. Thus a given exercise (test item) could be classified according to targeted Division and as well according to its range. A special ordered pair notation was used to convey this information: an item designated (II, I-III) was targeted at Division II but was included in the item battery for Divisions I, II and III. Such an item would be useful for tracing growth over grades 3 through 9. As well, it was assessing content encountered in the curriculum in grades 4-6.

From a conceptual/pedagogical point of view, the notion of target and range for common items permitted the identification of three types of items as follows:

1. Preview items. If an item was given at a Division lower than its target, it provided "preview" information on that content.

2. Target items. All items targeted at a given Division were given at that Division. Target items essentially provided information on how students performed on mathematical concepts and skills specified for their level by the Alberta Program of Studies.

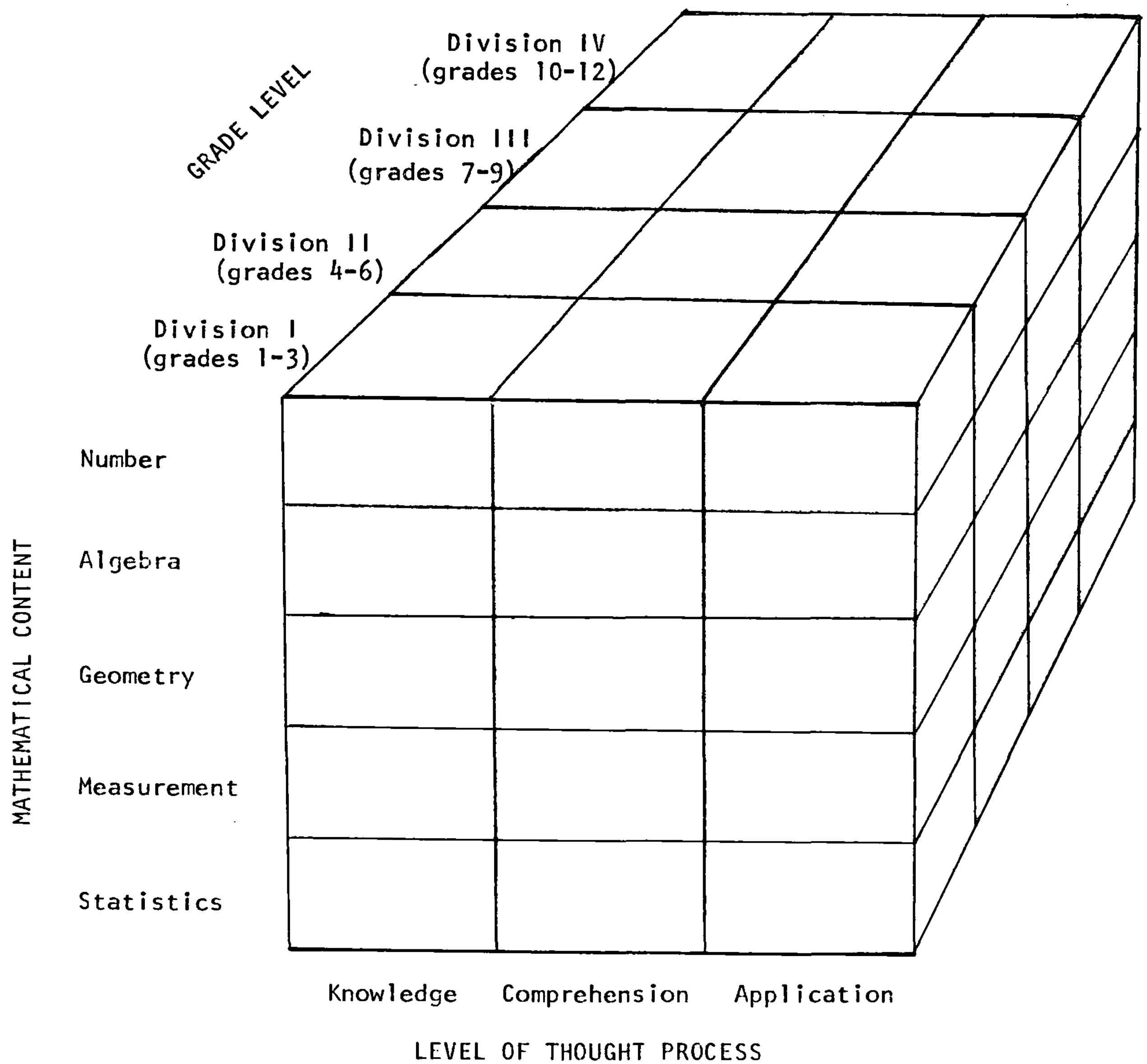

Figure I-1 Mathematical Content Specification Model.

3. Review items. If an item was given at a Division higher than its target, it provided "review" information on how students performed on material they should have mastered at least three years earlier.

The variables mentioned thus far all relate to ways of structuring the estimates of the index numbers. As well, and for its own sake, student background information of the following kinds was collected: (1) class time devoted to mathematics, (2) use of hand-held calculators, (3) attitude towards mathematics, and (4) usefulness of mathematics as perceived by students.

1.3.2 Methods of Data Collection

The complete battery of items given at a particular grade level (3, 6, 9, or 12) consisted of 120 exercises. The make-up of each battery is indicated in Figure 1-2 in which the number of items classified by content and targeted Division is shown. For example, the item battery for Division I contained 64 exercises on number; 48 of these were targeted at Division I, 11 at Division II, 5 at Division III, and 0 at Division IV. In fact, as shown in the figure, no items in the Division I battery were targeted at Division IV. On the other hand, six items in the Division IV battery were targeted at Division I, in keeping with the guideline that for purposes of assessment, review information is more valuable than preview information.

In selecting the 120 items for each Division, two rounds of screening by classroom teachers took place. In round 1 eight teachers, two each at grades 3, 6, 9, and 12, examined approximately 1500 test items selected from similar assessments done in Canada, U.S.A. and Great Britain. The criteria of clarity, fairness and representativeness were used by the teachers to rate each item. In round 2, eight other teachers from a rural area in Alberta examined the approximately 1200 items rated acceptable or better by the round 1 teachers. This second set of teachers assessed each item according to the range for which it would be useful. On the basis of the information provided by these two rounds of item validation, the contractors were able to select the 120 items for each Division battery. As well, the Steering Committee examined each item in detail and made comments for revision. Since many items were common to more than one battery only 275 different items were needed, these 275 being distributed over the four Divisions.

To assist in the data collection, the 120 items for each Division were packaged into booklets. At Divisions I, II, and III, the booklets contained twenty items each. At Division IV the booklets contained thirty items each. Figure 1-3 gives information on the format of the booklets at each Division. Over all, approximately 70% of the exercises were multiple choice in format, the remaining 30% being open-ended.

Division	Content	Target Division I	II	III	IV	Total
I	Number	49	10	5	0	64
	Algebra	9	2	1	0	12
	Geometry	20	2	4	0	26
	Measurement	10	2	0	0	12
	Statistics	4	1	1	0	6
	Total	92	17	11	0	120
II	Number	14	25	12	2	53
	Algebra	4	7	5	1	17
	Geometry	8	14	8	2	32
	Measurement	3	6	2	1	12
	Statistics	1	3	2	0	6
	Total	30	55	29	6	120
III	Number	3	11	15	6	35
	Algebra	4	7	21	9	41
	Geometry	2	7	11	6	26
	Measurement	2	3	5	2	12
	Statistics	1	1	3	1	6
	Total	12	29	55	24	120
IV	Number	1	1	8	8	18
	Algebra	3	6	12	27	48
	Geometry	1	2	5	10	18
	Measurement	1	1	3	7	12
	Statistics	0	1	2	3	6
	Consumer Math	—	—	—	18	18
	Total	6	11	30	73	120

Figure 1-2 Content Sampling Matrices per Division
(The entries refer to number of items)

Division	Booklet letter	No. of items	No. of open ended items	No. of multiple choice items	Separate answer sheet
I	A	20	20	—	No
	B	20	20	—	No
	C	20	—	20	No
	D	20	—	20	No
	E	20	—	20	No
	F	20	—	20	No
II	A	20	20	—	No
	B	20	20	—	No
	C	20	—	20	Yes
	D	20	—	20	Yes
	E	20	—	20	Yes
	F	20	—	20	Yes
III	A	20	13	7	No
	B	20	12	8	No
	C	20	—	20	Yes
	D	20	—	20	Yes
	E	20	—	20	Yes
	F	20	—	20	Yes
IV	A	30	27	3	No
	B	30	—	30	Yes
	C	30	—	30	Yes
	D	30	—	30	Yes

Figure I-3 Booklet Format per Division

Use of the booklet format was intended to minimize the amount of time any given student had to spend writing since only one booklet was completed by any one student, a task that required no more than forty minutes. Two rounds of pilot testing took place in the Edmonton area and revisions in the booklets were made on the basis of the information obtained.

The sample of students was selected by the Student Evaluation and Data Processing Branch of Alberta Education using the methods of probability sampling. In essence, every student attending schools supported by public funds had an equal chance of being selected. The sample consisted of 102 schools offering grade 3; 101 schools offering grade 6; 56 schools offering grade 9; and 50 schools offering grade 12. The technical details of the matrix sampling scheme are fully described in the technical report.

The test materials were packaged in classroom lots and mailed to the sample schools. Each classroom package contained approximately equal numbers of each booklet for the given grade level. The number of students writing each booklet at each grade level is shown in Figure I-4. The return rate from the schools was 100% - i.e. return packages were received from all of the schools in the sample. However, on the day of the testing (Wednesday, May 17, 1978) a small portion of students did not attend school. In comparing returns to the number of students in each class as supplied by principals, the rate of return from students was 95% at grade 3, 90% at grade 6, 93% at grade 9, and 73% at grade 12.

1.3.3 Methods of Analysis

Estimates for the index numbers were simply the proportion (percentage) of a given type of respondent giving the correct response to a given exercise. Types of respondents were defined by the personological/demographic variables listed under 1.3.1. Thus for an individual exercise, the estimate is the percentage of correct response over a given type of respondent. For sets of items grouped according to mathematical content specifications, the estimate is simply the arithmetic mean (average) of the single item estimates. No weighting scheme was used or needed to be used in calculating the arithmetic means since the number of items assigned to a given mathematical topic was in proportion to its emphasis in the curriculum at that Division, as specified by the various guides issued by Alberta Education for use by schools.

1.3.4 Utilization of Results

This study was done under contract for the Minister's Advisory Committee on Student Achievement (MACOSA). What MACOSA or the Minister himself will do with the results is not known by the contractors. What the contractors do know is that the results will be made public. It is for this reason that the notion of an index was dominant in the conception of this assessment. The expectation is that the present study is the beginning of something that might be called the

SAMPLE RESOLUTION

DIV I:	ALL	BOYS	GIRLS	S SCHOOL	M SCHOOL	L SCHOOL
BOOK A	563	282	281	149	245	169
BOOK B	573	299	270	157	248	168
BOOK C	570	294	265	159	248	163
BOOK D	566	281	282	150	255	161
BOOK E	548	300	247	144	243	161
BOOK F	553	270	278	147	244	162

DIV II:	ALL	BOYS	GIRLS	S SCHOOL	M SCHOOL	L SCHOOL
BOOK A	520	284	236	150	181	189
BOOK B	518	258	259	148	182	188
BOOK C	501	252	248	144	182	175
BOOK D	511	258	251	154	181	176
BOOK E	511	267	244	157	178	176
BOOK F	507	258	249	147	179	181

DIV III:	ALL	BOYS	GIRLS	S SCHOOL	M SCHOOL	L SCHOOL
BOOK A	542	270	271	96	143	303
BOOK B	543	271	272	101	142	300
BOOK C	522	271	247	96	132	294
BOOK D	531	261	268	96	139	296
BOOK E	525	251	269	94	140	291
BOOK F	531	276	251	93	139	299

DIV IV:	ALL	BOYS	GIRLS	S SCHOOL	M SCHOOL	L SCHOOL
BOOK A	877	424	422	210	228	439
BOOK B	855	406	449	201	227	427
BOOK C	857	405	450	206	226	425
BOOK D	855	411	439	207	219	429

Figure 1-4

Mathematics Proficiency Index (MPI) to be used and interpreted by a wide cross-section of the public in much the same way as the Consumer Price Index (CPI) is used and interpreted by anyone interested in the economy. The provision of index numbers by the present assessment should mark the beginning of a new level of awareness if government decides to make the MPI a reality.

1.3.5 Desired Precision

In studies of the present type it is desirable to know what sort of confidence can be placed on the estimates of the index numbers, i.e. on the percentage of success on an individual exercise. In particular, it is desirable to know whether there is a significant difference between any two percentages. The sample for the study was chosen so that any difference of 9 or more percentage points would be significant at the 0.01 level. For example, if grade 3 students had a 48% success rate on a given exercise whereas the grade 6 students performed at the 58% level, such a difference could be termed a real difference with 99% confidence. In like manner, the sampling scheme ensures that differences exceeding 6 percentage points can be said to be significant at the 0.05 level. The details of the sampling scheme that delivers the results with this precision are given in the technical report.

1.4 Time Line for the Study

The study was carried out in seven phases. These phases are briefly described below since they provide a useful way of summarizing the study as a whole.

1.4.1 Design (April - May, 1977)

Detailed plans regarding

1. sampling
2. mathematical content specifications
3. item validation procedures
4. test piloting procedures
5. data collection scheme
6. data analysis plan
7. final report

were drafted.

1.4.2 Item Specification and Collection (May - September, 1977)

A large bank of test items was collected and classified according to the content-levels matrix developed in the first phase. Released items from numerous state, provincial, national and international mathematics assessments served as the source.

1.4.3 Item Validation (September - December, 1977)

First a group of eight mathematics teachers, two from each of grades 3, 6, 9, and 12, examined all of the items that had been collected and classified according to the content-levels matrix in phase 2. The teachers examined each item for face validity, clarity, and relevance to the Alberta scene.

A second group of eight teachers took the items which had been screened by the first group and indicated at what grade levels it would be useful and feasible to use the item. In essence this group of teachers helped to identify which item would be useful for providing cross-sectional growth information.

1.4.4 Booklet Development and Pilot Testing (January - April, 1978)

Specifications were drawn up regarding the percentage distribution of items over the content areas (number, algebra, geometry, measurement, statistics, and consumer mathematics (Division IV only)) for each Division separately.

As well, specifications were drawn up for the common items indicating how many items would be given at more than one Division, which particular Divisions the item would span, at which Division the item would be targeted, and finally what content would be covered by the item. The details of the planning and the final solutions achieved are described in detail in the technical report.

The total number of items for all four Divisions was 275. Approximately half, 131 to be precise, were common items (given at more than one Division). The remaining 144 were unique items (given at only one Division). Each Division received 120 items, some unique, and some common. These 120 items were divided into six booklets of twenty items each at Divisions I, II, and III; at Division IV, four booklets of thirty were developed.

The booklets underwent two stages of pilot development before final arrangements were made regarding printing and administration procedures.

1.4.5 Data Collection (May, 1978)

All data collection was done via the mail with backup communication provided by long distance telephone. Test packages were mailed by May 5, 1978 to all of the schools in the sample. Postage paid return shipping envelopes were included with every package.

1.4.6 Data Analysis (June, 1978)

Approximately 25% of the items were open ended, the remaining 75% being multiple choice. At the Division II, III and IV levels, the multiple choice items required the use of specially prepared optical scoring answer sheets. The multiple choice items at Division I did not require the use of a separate answer sheet. Two experienced mathematics teachers were hired to score the non-optically-scored booklets. Special programs were written by Computer Services of The University of Alberta to carry out the analyses, the results of which are presented in Part II.

1.4.7 Report Writing (June - July, 1978)

A preliminary report was presented to the Steering Committee on June 21, 1978. Discussions at that time settled the details of the final report form.

1.5 Findings

1.5.1 Content

Number. The results on number showed a great deal of variability with performances ranging from a low of 11% on some items to a high of 99%.

At the Division I level strong performances were obtained on the fundamental skills of primary school mathematics -- performances ranging between 91% and 98% were given on items dealing with cardinal number of a set, reading 3-digit numbers, counting by tens, ordering 4-digit numbers, and the basic facts of addition and subtraction. Rather weak performances were obtained on exercises dealing with rational numbers -- performances ranged between 11% and 72%. At the low end were simple decimal tasks asking that 0.3 be read as three tenths, and at the high end were tasks dealing with recognizing common fractions.

At Division II the strong start made on the fundamental skills was continued with performances in the 90's given on exercises dealing with basic facts of multiplication and division, adding

and subtracting and multiplying multi-digit numbers, reading 5-digit numbers, and rounding to the nearest ten. As with Division I, the work with rational numbers was weak both in common fraction form or decimal form. As well work with integers gave low performances.

For Division III the strong areas for Divisions I and II remained strong. This was particularly true for whole number work. In addition the exercises on place value resulted in strong performances (typically 85%-95%). As well, performances in the 80's were given on whole number applications. Somewhat mediocre performances were given on percent and integers and as at the Division I and II levels, the results on common and decimal fractions failed to reach the level of whole number work although computing with decimal fractions rose significantly.

For Division IV the picture remained pretty well the same as for Division III. Division IV continued to do well on the exercises done well by Division III but failed to improve significantly on the weak areas of Division III.

Algebra. The algebra program receives little emphasis before Division III. The first two Divisions, however, showed some understanding of some fundamental algebraic concepts. Over 80% of the students in Division I showed an understanding of the use of the box "□" in a simple algebraic sentence while two-thirds of them could translate an English statement into a sentence containing a box. Perhaps it is encouraging that one-third indicate an understanding of the concept "more than" and the accompanying symbolism ">". By Division II, two-thirds of the students are using this idea on inequality correctly, while the last two Divisions show 90% correct usage.

The focus of the Division II algebra was verbal problems. Simple problems containing 3 or 4 sentences and one or two operations were answered correctly by less than 75% of the students. Division III shows only slightly better on these problems. Over half of Division II students can solve a simple linear equation involving "x" but all three of the first Divisions show up poorly in solving an equation when x is negative.

Division III does not show well on simple linear expressions and where the answer requires the use of a variable, they show up very poorly. On standard "age" problems Division III does well but any variation making the problem slightly unusual brings about marked drops in the scores. Only half of Division III students understand the linear function.

Division IV does well on simple algebra involving equality, inequalities, simple verbal problems, evaluating polynomials, and simultaneous linear equations. About 75% can interpret a formula involving a linear function. There seems to be several areas in mathematics that can be dealt with by half of the students: quadratics, graphs of linear equations, exponents, division of polynomials, and solving rational expressions. There are also several areas where students perform at less than 50% and indeed much lower: conics, composite functions, exponential

functions and trigonometry. The Division IV results are characterized by poorer responses to problems where answers are in terms of variables and depressed scores where problems involve something slightly different.

Geometry. Geometry, together with number and algebra, is a fundamental part of any mathematics program. The amount of emphasis that it receives probably varies greatly from school to school but it probably is included in all school mathematics programs from the early grades and on.

It seems safe to conclude that students do well enough on topics that receive emphasis in the schools. On the other hand they do less well on topics that are not emphasized. A case in point occurs with some exercises related to points, lines, planes and space. One of these exercises in two dimensions had percentages of correct responses of 80, 83 and 88 at Divisions II, III, and IV, respectively, while a similar exercise in three dimensions had percentages of 34 and 36 at Divisions III and IV. Other weak areas, probably because of a lack of exposure, were transformations and symmetry.

Performance levels were relatively high on the following topics: angles, triangles, circles and congruence. In contrast, the following topics were not handled nearly as well: similarity, line relationships (parallelism and perpendicularity), and polygons.

In a number of exercises students seemed to be misled by irrelevant perceptual cues, e.g., a square supported on one of its corners was not recognized by a significant percentage of students. The application exercises, many of which required high verbal decoding skills, were difficult for many students as well.

Measurement. Most of the measurement exercises were simple Knowledge items which tested familiarity with the metric system or the application of metric units in geometric settings. The telling of time, figuring with money, and measuring length had strong beginnings in Division I and rose to 90% performances by Division II. Questions on perimeter, area and volume fared less well resulting in mediocre performances even at Division IV (70%). In particular, exercises dealing with interrelationships among length, area and volume were very poorly handled (30-40%) at Divisions III and IV. As well, items dealing with interrelationships among metric prefixes showed that over one quarter of grade 12 students aren't familiar with them.

Statistics. All of the items in statistics were at the Comprehension or Application level. Division I scored high on picture graphs, less well on bar graphs and about 40% on line graphs with only one-third being able to construct a bar graph. Division II does well on the bar and line graph and 50% on the circle graph. A problem based on a multiple-line line graph showed half of the students as being able to do it. Division III follows the trend with almost 80% responding correctly to the circle graph. Problems based on graphs bring in slightly lower

performances. Interpreting a line graph showed high levels of performance from both Division III and IV. The latter Division did poorly on a simple probability question and very poorly on a problem using elementary statistical concepts. In all cases orderly and consistent growth across Divisions was noted.

Consumer Mathematics. The inclusion of eighteen exercises skills in consumer-related problems was a unique feature of the Division IV. battery. The exercises were all classified as Applications of mathematics.

The performances were not high with a range of percentages of correct responses from 2 to 85 with a mean of 51%.

Consumer related mathematics is not emphasized in the high school mathematics program. Consequently performance levels were not high, particularly on items that required a modicum of knowledge about specialized terminology e.g. margin of profit, mill rate, etc. This part of the assessment was recognized as not being related to the curriculum but it was argued that these consumer-related skills and knowledge are required of adults and so it was desirable to know how well grade 12 students fare.

Performance levels were highest on the following topics: comparative shopping, interest calculations, consumer credit, payroll, taxation, and discount computations. These are probably the most prevalent consumer related problems that people must face.

In contrast, performance levels on the following topics were not as high: profit and loss, banking, credit card statements, insurance, investments, and real estate.

1.5.2 Other Variables

Sex. Over the years it has become the accepted belief that boys achieve better than girls in mathematics. The results of the present study show conclusively and unambiguously that the old belief is valid for Alberta in 1978. Not only did the boys do better on knowledge items but the difference between sexes became stronger for the Comprehension and Application items. Furthermore the sex difference became more pronounced with each increasing grade level resulting in supplements of 5 percent or more in favor of boys at grade 12.

School Size. At each Division, school were classified as small, medium and large. As with sex, school size showed significantly in accounting for achievement differences. Interestingly, the differences interacted with grade level. For Division I, the small school (enrolment under 30), showed the best and quite consistently so over the different content areas. For Divisions II the

medium school (enrolment between 31 and 70) began to emerge as dominant and in Division III the medium school (enrolment between 41 and 101) showed clear dominance. Somewhat surprisingly, school size failed to make any difference at the Division IV level. Perhaps surprising was the fact that the large elementary school (enrolment over 70) was quite consistently the poorest performer. Since most of the large elementary schools occur in centres of large population it is essentially the city elementary schools that performed less well than the small (and usually rural) counterpart.

Grade 12 Mathematics Program Variable. At grade 12 the purpose of the assessment was to measure general proficiency in mathematics. This implied that the assessment not be tied to a particular mathematics curriculum at grade 12.

Although the assessment was to measure general proficiency it was obvious that performances would still vary as a result of differences in mathematical programs. Consequently student mathematical programs was used as a variable in the analysis of the data. The mathematical programs were defined in terms of the following six alternatives: one mathematics course, Math 20, Math 23, Math 25, Math 30, and Math 33.

As expected the highest performance levels came from the Math 30 sample. Next highest came from Math 20 and Math 33. Finally, the three remaining alternatives did least well, i.e., Math 23, Math 25, and one mathematics course.

The greatest differences in performance across mathematical programs occurred on exercises related to number and algebra. This result seems reasonable because it is in these areas where the Math 30 student receives the greatest preparation.

Lesser differences in performance occurred on exercises related to geometry, to measurement, to statistics and to consumer mathematics. These results probably reflect less difference in mathematical preparation on these topics.

Student Background Information. Our survey of student backgrounds showed that student use of calculators outside the classroom is far higher than inside the classroom. Only at Division IV is the amount of usage getting close.

At grade 3 one in three students find mathematics their favorite subject while this is reduced to one in seventeen at grade 12. And while most students think mathematics is valuable, only 33% of grade 12 students find it "very valuable". Such a trend seems almost to be inevitable in that the grade 12 student is becoming exposed to more attractions.

The class time spent on mathematics is quite variable at Division I and II ranging from 35 to 60 minutes per day. Division III has a smaller range with most schools at either 40 or 50 minutes. On the average the amount of time spent on mathematics decreases significantly from Division II to Division III.

1.6 Recommendations to MACOSA

1.6.1 Main Recommendations

The Contractors strongly recommend that:

1. Alberta Education take immediate steps to make the assessment an on-going endeavour and in this way establish a Mathematics Proficiency Index (MPI) for Alberta.

2. Alberta Education develop a set of terms of reference for the MPI through the conduct of a needs assessment study dealing with mathematics proficiency in Alberta.

3. Access to all the reports written by the contractors be made available to all bodies making curricular decisions in Alberta. In particular the Condensed Report should be given as wide circulation as possible. Interpretative panels should be set up with a view to making the information in the reports more easily understood by a wider variety of interested people.

4. Computerized item banking be developed and maintained.

5. All the data gathered by the contractors be made available to any bonafide researcher interested in the information providing anonymity of students and schools be protected.

6. The focus of the Division IV test battery be more on "school leaving" competencies rather than the various senior high school mathematics programs.

7. The variable of sex be subjected to further study with a view to understanding the reasons for the lower performance of girls in Alberta. Such study should result in short range and long range plans to assure that girls have an equal opportunity to acquire competence in mathematics, a competence, the absence of which, acts in many explicit as well as subtle ways in preventing women from having the same freedom as men in determining their occupational as well as intellectual destiny.

8. The variable of school size be examined in a much more explicit and controlled fashion than was possible in a survey. The results of the present study suggest that the large school, particularly at the lower grade levels, may not be as effective an organization for delivering mathematics instruction as smaller units. This general issue should be subjected to

an intensive observational assessment to determine the reasons for the differential effect of school size.

1.6.2 Procedural Recommendations

9. Procedures for developing additional items should be undertaken as early as possible. The need for new items is especially felt at Division I.

10. The common items that are used at more than one Division should be continued.

11. The number of items used at each Division should be increased from 120 to 150. The use of open-ended items for certain types of questions should also be continued.

12. The teacher validation-of-items procedures that were used were very effective and should be made a part of any future assessment.

13. The test administration procedures used were adequate and should be used for the next assessment.

14. The background variables -- attitude, value of mathematics, usefulness of mathematics -- and the "other variables" -- sex, size of school, and mathematics program at grade 12 -- should be included in the next assessment.

1.6.3 Background to the Recommendations

Recommendation 1 develops out of the original statement of the problem. Greater understanding of the effects of the mathematics curriculum will be achieved as we examine how these effects change over time. The Mathematics Proficiency Index would use the current assessment results as baseline data. The actual index could consist of different indices for the various curriculum strands or could be reduced through appropriate weightings to a single index. Recommendation 4 follows as a consequence of this recommendation since it is essential to use a large portion of identical items from one assessment to the next. Computerized item banking is extremely convenient for information across many Divisions on a particular item.

Recommendation 2 arises from a concern about the usage of the information. If knowledge were available as to who was to use the information, the whole structure of the item battery could be focussed appropriately. In particular, Recommendation 6 reflects a concern about the mathematical competence of the early adult. Knowledge of "school leaving" competencies can have implications for the structure of the mathematics curriculum at all levels.

Recommendation 3 is made from a belief that the information collected is very valuable to classroom teachers and every effort should be made to encourage them to make use of it. Interpretative panels are at liberty to make judgements based on their own points of view. These judgements are valuable but should not be confused with the information provided in the assessment reports.

Recommendation 5 suggests that Alberta Education be the repository of all the data collected. Many interesting studies of the data are possible beyond the current study. Alberta Education should encourage researchers to engage in such studies.

Recommendation 7 is concerned with the possibility that girls may not be treated fairly in Alberta in terms of equal opportunity to learn mathematics. It has become clear in recent years that girls in the U. S. are at a disadvantage in mathematics due in large part to reasons beyond their control. For example, it has been found that teacher referrals (Gregory, 1976) as well as standardized achievement tests (Tittle, 1974) are sex biased. Such bias should be studied in Alberta to determine its nature and extent. Indeed, since the mandate of the present study called for the use of existing test items insofar as they were appropriate to the provincial course of studies, over 90% of the items were selected from assessment studies conducted in recent years in the U. S. Thus the bias that Tittle (1974) found in standardized achievement tests may also pervade test items used in American assessments. The possible bias present in the present study is thus evident. Further study of sex bias in mathematics assessment is thus highly recommended, not only to control for it in assessments of this kind, but to fully document the possible unfair treatment that girls may be subjected to by the school systems of Alberta.

Recommendation 8 is concerned with the differences concommitant with large school size which seem to detract from mathematics learning. Discovering what these concommitant differences are could result in programs and organizational changes that might produce improved learning in all schools.

Recommendations 9 through 11 refer to items to be used in a future assessment. Using the item sampling design as used in the current study, the number of items could be increased 5 per booklet without any modification in the administration of the tests. While it is clear that the use of common items reduced the amount of target information, the information gained by a common item, while not related specifically to the curriculum, enables one to get a total picture of mathematics learning. It provides a view of the development of mathematical concepts through the Divisions. Open-ended items, although more costly to grade, are essential for assessing basic facts and computation and can effectively be used in many other questions.

Recommendations 12 and 13 grow out of subjective judgements of the researchers. No assessment will work without full cooperation from classroom teachers. That cooperation can be secured through teachers having confidence in the study and knowledge about its operation in the broadest sense. The validation and administration procedures were designed to maximize these two aspects.

The variables named in Recommendation 14 provide not only information directly relevant to this assessment but they also make accessible, at very low cost, answers to other important research questions. It is also possible that depending on the "needs assessment" other background variables might be of interest in a future assessment.

DATE DUE			